



Janssen, L. M. M., Drost, R. M. W. A., Paulus, A. T. G., Garfield, K., Hollingworth, W., Noble, S., Thorn, J. C., Pokhilenko, I., & Evers, S. M. A. A. (2021). Aspects and Challenges of Resource Use Measurement in Health Economics: Towards a Comprehensive Measurement Framework. *PharmacoEconomics*, 39(9), 983-993. <https://doi.org/10.1007/s40273-021-01048-z>

Publisher's PDF, also known as Version of record

License (if available):
CC BY-NC

Link to published version (if available):
[10.1007/s40273-021-01048-z](https://doi.org/10.1007/s40273-021-01048-z)

[Link to publication record in Explore Bristol Research](#)
PDF-document

This is the final published version of the article (version of record). It first appeared online via Springer at <https://doi.org/10.1007/s40273-021-01048-z>. Please refer to any applicable terms of use of the publisher.

University of Bristol - Explore Bristol Research

General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available: <http://www.bristol.ac.uk/red/research-policy/pure/user-guides/ebr-terms/>



Aspects and Challenges of Resource Use Measurement in Health Economics: Towards a Comprehensive Measurement Framework

Luca M. M. Janssen¹ · Ruben M. W. A. Drost¹ · Aggie T. G. Paulus¹ · Kirsty Garfield² · William Hollingworth² · Sian Noble² · Joanna C. Thorn² · Irina Pokhilenko¹ · Silvia M. A. A. Evers^{1,3}

Accepted: 26 May 2021
© The Author(s) 2021

Abstract

Background While the methods for conducting health economics research in general are improving, current guidelines provide limited guidance regarding resource use measurement (RUM). Consequently, a variety of methods exists, yet there is no overview of aspects to consider when deciding on the most appropriate RUM methodology. Therefore, this study aims to (1) identify and categorize existing knowledge regarding aspects of RUM, and (2) develop a framework that provides a comprehensive overview of methodological aspects regarding RUM.

Methods Relevant articles were identified by enrolling a search string in six databases and handsearching the DIRUM database. Included articles were descriptively reviewed and served as input for a comprehensive framework. Health economics experts were involved during the process to establish the framework's face validity.

Results Forty articles were included in the scoping review. The RUM framework consists of four methodological RUM domains: 'Whom to measure', addressing whom to ask and whom to measure; 'How to measure', addressing the different approaches of measurement; 'How often to measure', addressing recall period and measurement patterns; and 'Additional considerations', which covers additional aspects that are essential for further refining the methodologies for measurement. Evidence retrieved from the scoping review was categorized according to these domains.

Conclusion This study clustered the aspects of RUM methodology in health economics into a comprehensive framework. The results may guide health economists in their decision making regarding the selection of appropriate RUM methods and developing instruments for RUM. Furthermore, policy makers may use these findings to review study results from an evidence-based perspective.

1 Introduction

As healthcare needs increasingly outweigh available resources, policy makers are more commonly using economic evaluations to guide efficient resource allocation [1]. As part of such evaluations, the total costs associated with

an intervention are calculated by multiplying resource use measurement (RUM) estimates (e.g. the number of general practitioner visits) by the corresponding unit prices (e.g. €33 per visit) [2]. Accordingly, measuring the true quantities of the resources utilized is a vital part of generating valid costing estimates [3, 4]. In economic evaluations conducted from a societal perspective, i.e. when all relevant costs and benefits are accounted for regardless of where they occur, RUM covers a broad spectrum of services in the healthcare sector (e.g. general practitioner visits) as well as in other sectors [5]. These may include sectors such as the social care sector (e.g. home carer visits), the education sector (e.g. special education services), and the criminal justice sector (e.g. contacts with the police) [6]. In addition to services, time and money are also considered resources; therefore productivity losses, time spent providing informal care, and out-of-pocket expenses associated with one's health may also be covered in RUM. The variety of resources to be measured in health

✉ Luca M. M. Janssen
luca.janssen@maastrichtuniversity.nl

¹ Department of Health Services Research, Care and Public Health Research Institute (CAPHRI), Faculty of Health Medicine and Life Sciences, Maastricht University, P.O. Box 616, 6200 MD Maastricht, Maastricht, The Netherlands

² Department of Population Health Sciences, University of Bristol, Bristol, UK

³ Trimbos Institute, Netherlands Institute of Mental Health and Addiction, Centre for Economic Evaluation and Machine Learning, Utrecht, The Netherlands

Key Points for Decision Makers

Resource use measurement (RUM) is often based on practicality rather than on evidence-based methodologies. This could result in inaccurate results of health economic studies and could inherently misinform policy decision makers. While some methodological studies have focused on specific aspects of RUM, a general overview that constitutes RUM as a whole is lacking.

This study provides a framework of RUM aspects and domains, including the current recommendations based on literature review and expert input. This could help health economics researchers select the best suitable measurement approach for their study and thus improve the chances to obtain accurate outcomes.

RUM is complex and combining existing recommendations that focused on one aspect generated new insights. In addition, it exposes the research gap of where methodological evidence is lacking most and future methodological research could focus on. It stimulates further research to use evidence-based RUM approaches, thus providing improved evidence for policy decision making.

economics research calls for a harmonized evidence-based approach towards RUM.

RUM is part of the bigger cost measurement process in health economics research [7]. The process of cost measurement is often divided into a three-step process, containing the (1) identification of costs, (2) measurement of costs, and (3) valuation of costs. During the first step, identification, the costs that are considered relevant for inclusion for a study are determined. This depends on, among other factors, the aim of the study, the perspective, and the target audience of the results (e.g. policy makers, hospital managers). Recommended perspectives differ per country [8] and some countries may have no national guidelines at all. Identification is finalized before the measurement of costs (step 2). The second step refers to the actual measurement of resources used and is referred to as RUM. The purpose of the second step is to quantify the total resource use in a given time span per individual or per population, depending on the aim of the study. The final step in the process is to determine the unit costs per unit of measurement to be able to calculate the total costs. These results can then be used to calculate the total burden of a disease or can serve as part of cost-effectiveness analyses.

RUM, the second step of the process, is known to be a challenging and time-consuming, yet essential, step in economic evaluations [3, 9]. However, literature and guidance

on the appropriate methods for RUM remain scarce as the existing guidelines focus predominantly on resource use valuation and outcome measurement [3, 10, 11]. A review on current methodological guidelines in 33 countries addressed only the types of costs that could be relevant for inclusion in economic evaluations, but did not provide guidance on how to appropriately measure those [12]. Current health technology assessment (HTA) guidelines worldwide (e.g. in the US, the UK, Australia and The Netherlands [2, 13–15]) lack information on adequate measurement of resources. Its current focus is predominantly on cost valuation and on quality-of-life measurement, but fails to provide concrete guidance for health economists on RUM aspects, i.e. how to obtain resource use data, how to define the most optimal recall period, or how to select the most appropriate RUM instrument, among others. Consequently, a wide variety of measurement methods exist and the choice of a measurement method often relies on practicality rather than on evidence [3], for example when a convenient recall period is chosen, rather than investigating which recall period leads to the most valid resource use estimates. Existing RUM instruments lack transparency in their development process; thus, it remains unclear which decisions were made with regard to their RUM approach [9]. This can affect the methodological quality of an economic evaluation and compromise the likelihood of acquiring accurate cost estimates [16]. Furthermore, the variability in national RUM guidelines on the recommended perspective [8] increases the variability of costing research, which compromises its comparability. Given that the use of different RUM methods may generate different results [17], more information is needed on the consequences of RUM methods.

Published methodological studies focus on specific questions related to RUM, such as, among others, ‘whom to ask’ or ‘do pharmaceutical data and patient-collected data agree?’ [18, 19]. Often, the generalizability of the results is limited and findings may be contradictory; hence, which recommendations to follow remains unclear. Until now, no framework is available that provides an overview of RUM recommendations, although this could provide guidance to researchers regarding evidence-based decision making in relation to RUM methodology. Therefore, this study aims to (1) identify and categorize published evidence from studies regarding the methodological aspects of RUM, and (2) develop a framework that provides a comprehensive overview of these aspects of RUM in health economics research.

2 Methods

This study was conducted in three chronological phases, as visualized in Fig. 1.

Phase 1 consisted of a workgroup meeting followed by a brief literature check. The purpose of the workgroup meeting was to brainstorm about the outline of the study. A workgroup consisting of seven health economists (LJ, SE, AP, CD, WH, SN, JT) brainstormed to identify existing RUM aspects in health economics, i.e. all concepts that are part of the complex phenomenon of RUM. Afterwards, the four methodological studies [3, 4, 20, 21] on RUM, familiar with the workgroup, were read full-text by one researcher (LJ) to validate the identified RUM aspects and to complement the list with other relevant RUM aspects. The RUM aspects identified during the expert meeting and the RUM aspects identified in the literature provided the input for the draft framework, developed in the next phase.

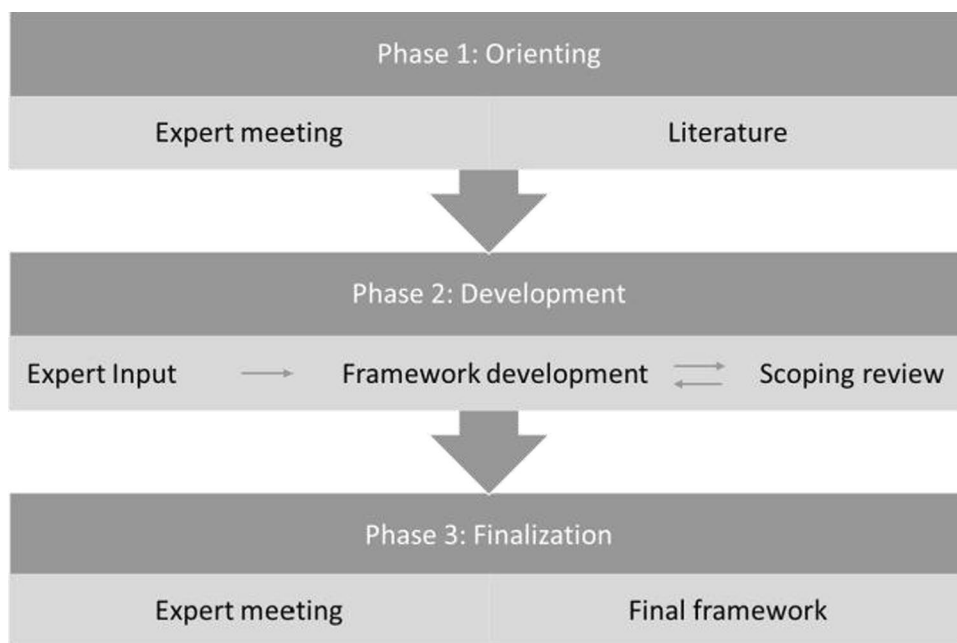
In Phase 2, the listed RUM aspects (that were identified in Phase 1) that seem to be connected, were clustered into broader domains. This clustering resulted in the draft framework, which was checked for validity by the PECUNIA consortium in a joint online meeting. This consortium consists of a group of health economics experts from 10 academic institutions in Europe [6]. The consortium studied the framework for face validity and completeness. No additional RUM aspects were mentioned and no other changes were suggested; ergo, the draft framework remained unchanged.

Thereupon, a scoping review was conducted that aimed to identify existing evidence-based findings and possible recommendations regarding the RUM aspects included in the framework. Scoping reviews are considered a suitable method to gather literature for lesser known fields, as their research questions are broader than those of literature reviews [22]. Scoping reviews aim to “map rapidly the key concepts underpinning a research area and the main sources

and types of evidence available” [23]. They can be useful in complex areas or areas that have not yet been comprehensively reviewed. Relevant literature for the scoping review was gathered using three approaches. First, a structured literature search was conducted. The search scopes of four existing systematic reviews addressing RUM that were identified by the experts in Phase 1 were used as a basis for the current literature search [3, 4, 20, 21]. Databases were chosen based on the scope and topic coverage and, as RUM is a relatively new and underexplored area of research, multiple databases were chosen. Six electronic databases, including EconLIT, EMBASE (Ovid), Education Resources Information Centre (ERIC), MEDLINE (PubMed), PsycINFO, and the Social Science Citation Index (SSCI; Web of Science), were searched in April 2018. The search strategy was developed to identify methodological papers addressing RUM recommendations (electronic supplementary material [ESM] 1). All papers were screened by one researcher (LJ) and, based on title and abstract, categorized as included, excluded or undecided. The articles that were categorized as undecided were screened by a second researcher (KG) and were included if the second researcher favoured their inclusion.

Second, the Database of Instruments for Resource Use Measurement (DIRUM), an open access database for resource use questionnaires, was hand-searched for relevant methodological papers in May 2018 [24]. At the time, DIRUM contained 97 articles and all were hand-searched, i.e. all titles and abstracts were read. Third, the members of the PECUNIA consortium (i.e. the same health economics experts who validated the framework in Phase 2) were asked to send potentially relevant articles to the authors.

Fig. 1 Methodological phases (1–3) of this study



The articles retrieved from both the structured literature search and the DIRUM database after title and abstract screening, and the articles that were recommended by the experts, underwent the same descriptive reviewing procedure. All articles were read full-text and were included in the study if the eligibility criteria were fulfilled. Methodological papers, i.e. studies that set out to answer a methodological question, that addressed RUM and were available in the English language were eligible for inclusion. The included articles were descriptively reviewed and the information regarding RUM aspects was extracted according to the data extraction template developed by the authors serving the aim of this study (ESM 2). The template was developed over online author meetings based on what authors believed was key information to be extracted. The results of the scoping review generated new insights and were used to update the framework in order to increase the mutual exclusivity of the aspects while balancing conciseness and completeness.

In the final phase, Phase 3, the framework was operationalized in multiple group discussions with the authors. During the operationalization process, definitions were created for the RUM domains and aspects to increase the usability of the framework for future studies. Developing definitions furthermore served to increase disambiguation of aspects and to decrease overlap as much as possible. Additionally, this step served as a final check whether the authors all agreed on the conceptualization of the framework. The framework was finalized and the results of the scoping review were categorized within it accordingly.

3 Results

The face-to-face expert meeting and four systematic reviews addressing RUM [3, 4, 20, 21] provided the input for the draft framework. Clustering of the RUM aspects resulted in a framework with six main methodological RUM domains and corresponding aspects (ESM 3): (1) What to measure? (2) Whom to ask? (3) How to measure? (4) How often to measure? (5) How to ensure the validity and reliability of RUM? (6) Additional considerations. This version was used as the outline for the scoping review. Figure 2 displays the flow of studies included after identification and screening, in the form of a Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) chart [25]. The search initially identified 7438 unique studies, 7393 of which were excluded after title and abstract screening, resulting in 45 remaining studies. Studies provided by the experts ($n = 48$) and studies identified via a search of DIRUM ($n = 35$) were added for full-text screening. Of the 117 unique studies included for full-text screening, 74 were excluded because they either did not focus on RUM ($n = 43$), were not a methodological study ($n = 33$), or were written in a language other than English ($n = 1$), resulting in

the inclusion of 40 studies that fulfilled the inclusion criteria (ESM 4); all were fully analyzed.

Data analysis generated new insights and led to a rearrangement of RUM aspects, resulting in the final framework (Fig. 3) that contains 4 RUM domains and 11 more specific RUM aspects. The RUM domain ‘How to ensure the validity and reliability of RUM?’ was renamed ‘psychometric quality’ and was merged as an aspect of RUM Domain 6 ‘Additional considerations’. Further changes are described in the footnote in the figure (ESM 3).

3.1 Resource Use Measurement (RUM) Domain 1: Whom to Ask

RUM Domain 1, ‘Whom to ask’, was discussed in 10 articles. This domain refers to the individual who consumes or uses a resource. Some personal characteristics (e.g. age) might influence the ability to recollect, and for some diseases (e.g. severe mental health illnesses), persons other than the individual him/herself might be better at recalling resource use. Furthermore, capturing only the individuals’ resource use (i.e. excluding the resource use of relatives) may not be sufficient to obtain relevant societal outputs. Therefore, this domain consists of two aspects—whom to ask and whom to measure.

3.1.1 Whom to Ask

Studies disagree on the effect of personal characteristics on the accuracy of self-reported RUM. Two studies report that differences in sex and age do not affect the ability to accurately recall [19, 26], while two other studies oppose this and conclude that patient characteristics such as age, health, and the amount of resource use may influence the likelihood of over- or underreporting [27, 28]. Another study added that educational attainment and functional ability may also affect the consistency of self-reports [29]. Literature seems to agree that if the target group includes people whose use of resources is high, recall problems are more likely to occur [27, 30, 31]. The studies also gave explanations for this phenomenon: people whose use of resources is high may be more severely ill and their ill health may directly affect their ability to recollect; moreover, people whose resource use is high have more to remember, as they have more resources to recall. These explanations also link to the RUM aspect ‘recall period’ in RUM Domain 4 ‘How often to measure?’. Furthermore, it may be possible that people (e.g. informal caregivers, parents) other than the patient him/herself are better at recalling resource use. Two studies report that parents can be good proxies for children, even though their recall is not perfect [30, 32]. We did not find any studies that oppose this finding, nor did we find any studies that recommend involving other proxies.

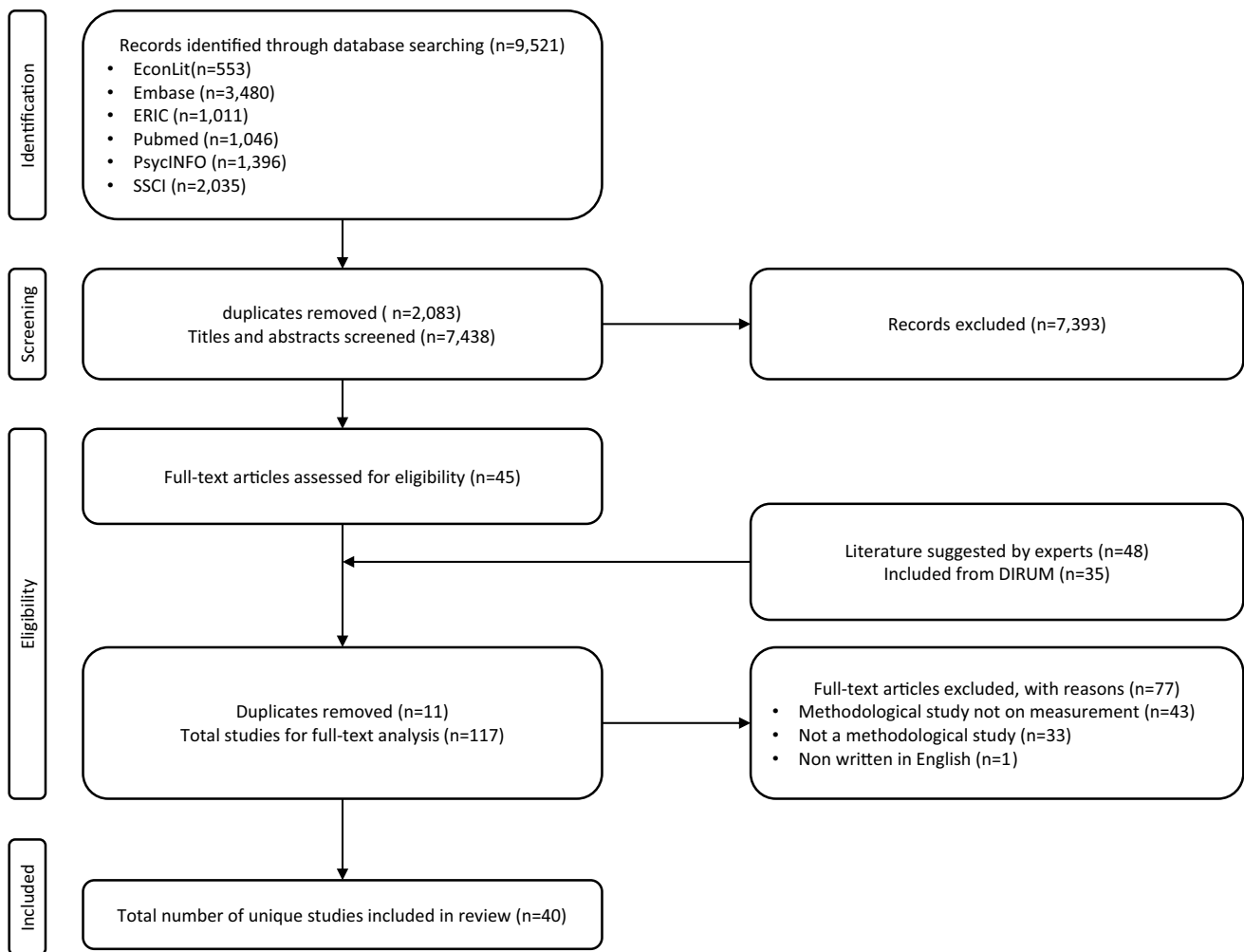
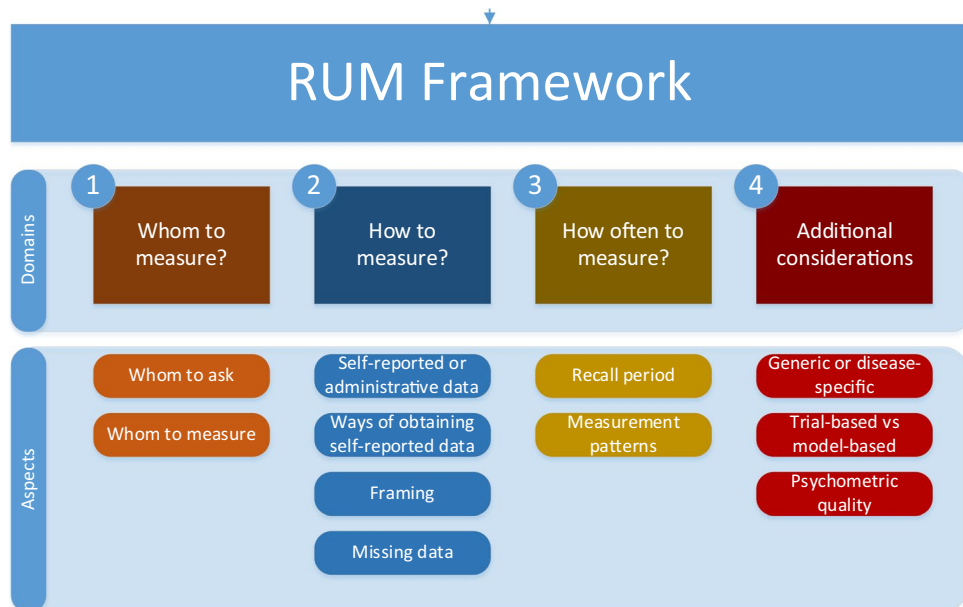


Fig. 2 PRISMA flowchart. *PRISMA* Preferred Reporting Items for Systematic Reviews and Meta-Analyses, *SSCI* Social Science Citation Index, *DIRUM* Database of Instruments for Resource Use Measurement

Fig. 3 Final Resource Use Measurement (RUM) framework



3.1.2 Whom to Measure

It could be argued that if economic evaluations are conducted from a societal perspective, the costs and benefits of people other than the research participant should also be included, to capture the total impact of a disease. This would imply that the unit of analysis, i.e. whom to measure, is broader than the individual's resource use alone. However, we did not find any recommendations regarding the appropriate unit of analysis in the reviewed literature.

3.2 RUM Domain 2: How to Measure

RUM Domain 2, 'How to measure', was discussed in 29 articles. The domain consists of four aspects: the type of data used (either self-reported or administrative data), the different ways that self-reported resource use can be obtained (e.g. online versus pen and paper questionnaires, conducting an interview to collect resource use information, or with the use of props such as a resource use diary), the way questions and answer options are framed, and the relationship between RUM and missing data. The latter addresses only the prevalence of missing data. Dealing with missing data during analysis is considered outside the scope of the current study.

3.2.1 Self-Reported or Administrative Data

The differences between self-reported and administrative data have been widely discussed in the literature. The level of agreement varies per study and depends on the type of resource use measured [3, 4, 17–19, 27–29, 31, 33–43]. One study reported that even if there is good agreement on resource use, this does not always lead to good agreement on costs [33]. For resource use items with high unit costs, a small difference in the frequency of using this resource may cause great variation in total costs and vice versa. It seems that in general, the preference for self-reported or administrative data depends on the perspective taken. Self-reported data on resource use may be preferred when more cost categories are included (e.g. in a societal analysis). As it is impossible to retrieve data on out-of-pocket expenses and costs outside the healthcare sector from data on medical records, it may be easier to use one questionnaire to collect all information, instead of combining data from multiple sources.

3.2.2 Ways of Obtaining Self-Reported Data

There are different measurement methods for obtaining self-reported data, and every method comes with specific advantages and disadvantages. Studies acknowledge an interaction between a questionnaire, a respondent and an interviewer

[44]. On the one hand, an interview setting may provide for an increase in patient recall when interviewers probe patients [18]. On the other hand, a more anonymous setting (without a researcher present) may increase item response and leads to more accurate reporting, as the presence of an interviewer may cause social desirability bias, i.e. the overreporting of socially desirable behaviour [44]. Furthermore, the interviewer could unintentionally increase task difficulty (for example when talking too fast), which in turn could decrease the efforts of the participants in formulating responses [45].

One study concluded that in general, people prefer face-to-face interviews over telephone interviews, and electronic modes of administration over pen and paper modes [44]. However, the mode of questionnaire administration chosen for the study should also be suitable for the target group, as relying on internet-collected data alone could potentially underrepresent the elderly or less educated [46].

3.2.3 Framing

Resource use questionnaires assume that respondents share the same meaning for terms; however, cultural and social language differences can influence the interpretations of questions and answer options [32]. It is therefore necessary to validate the RUM instrument and check if participants share the same meaning of the items. In addition, the response-choice order may affect the results. One study found that, if the options are presented visually, the respondent is more likely to select the first option in comparison with the subsequent options, while if the results are presented orally, the last option is more likely to be selected [44].

3.2.4 Missing Data

The mode of questionnaire administration and respondent characteristics may influence the likelihood of non-response. In terms of missing data, face-to-face administration is preferred over telephone or postal questionnaire administration [44, 47]. The higher the non-response, the higher the chance that missing data are missing not at random but depending on certain patient characteristics [44]. Face-to-face interviews have been identified as having the highest response rate [44]. As touched upon in Domain 2 'Whom to ask', persons with high resource use may experience survey fatigue, which can only be minimized by shortening the questionnaire [48]. Several general solutions were mentioned to prevent missing data, including (1) sending reminders, (2) sponsorship by an official or respected body, (3) online questionnaires, (4) using an interviewer, and (5) using a resource use log [44, 45, 49]. Efforts to prevent missing data should be concentrated around the main cost drivers

[50], as missing data on these resources has the most impact on the study results. Researchers could also focus on a narrower range of costs, to limit the patient burden. Missing data is thus linked to many other domains and aspects (e.g. ‘Whom to ask’ and ‘How often to measure’).

3.3 RUM Domain 3: How Often to Measure

RUM Domain 3 ‘How often to measure’ was discussed in 10 articles. The issue consists of two aspects: the recall period and the measurement pattern. The recall period encompasses the ideal recall period for different resources, and the measurement pattern refers to the follow-up period of RUM.

3.3.1 Recall Period

The ideal recall period is dependent on several aspects. While some studies were able to advise an ‘ideal recall period’ for specific types of resource use, other studies suggest that there is no general ideal recall period [26, 51–54]. Furthermore, the ability to recall may depend on the type of resource use, as evidence suggests that people have better recall of salient episodes (such as an overnight stay in a hospital) than of routine visits [30, 36, 55]. The recall period should be long enough to cover all types of resource use; however, a (too) long recall period could decrease the accuracy of the responses [26, 28, 33]. The higher the number of visits an individual needs to recall, the greater the margin of potential variability [28]. The need to recall considerable resource use may lead to both underreporting (i.e. participants are more likely to forget or to be unwilling to write everything down) and overreporting (i.e. participants might not remember the exact date of a visit and therefore may include visits outside the recall period) [33]. This would favour a narrower (i.e. shorter) recall period; however, a recall period that is too narrow can lead to leaving out infrequent but expensive events, and shortening a recall period cannot fully prevent recall bias [53, 56].

3.3.2 Measurement Patterns

There is a general need to minimize patient burden, and intermittent data collection may be used to address this need. However, this requires inter- or extrapolation to fill in the gaps, and average estimates from intermittent data collection may reduce the different patterns of resource use (i.e. a specific type or amount of resource use in a given time span). This could be important for diseases with seasonal differences (e.g. influenza), or when trying to identify patterns of resource use. On the one hand, intermittent data collection would require the participants to recall resource use over a longer period, if the participant is asked less often and covers a longer recall period to lower the patient burden. On

the other hand, asking less often but retaining a short time period and thus resulting in gaps that have to be filled by extrapolation would cause increased variability [57]. Ignoring resource use patterns with data extrapolation may cause inaccurate results. To overcome most of these issues, it is essential that the sample size is large enough when using intermittent data collection [58].

In conclusion, the ‘How often to measure’ domain comes down to the trade-off between limiting information loss (because of a recall period that is too narrow), recall error (because of a recall period that is too long) and patient burden (when measuring resource use too frequently, without an increase in data quality) [51, 53].

3.4 RUM Domain 4: Additional Considerations

RUM Domain 4, ‘Additional considerations’, was discussed in eight articles. This domain covers additional aspects that are essential for further refining the methodologies for measurement. The domain consists of three aspects: generic or disease-specific resource use, trial-based versus model-based economic evaluations, and the psychometric quality of RUM.

3.4.1 Generic or Disease-Specific

One study mentioned that a distinction should be made between general health loss or loss due to a specific disease [52]. Accordingly, the study suggested always asking participants whether another significant health problem other than the disease in the domain has had an impact on their use of resources. However, comorbidities and resource use that are an indirect consequence of one’s health increase the difficulty for respondents to make a clear distinction between generic resource use and disease-specific resource use. These indirect consequences may occur in the healthcare sector (physical pain as a consequence of mental illness) and outside the healthcare sector (productivity losses, costs in the education or criminal justice sector). Capturing all health-related resource use would reduce the risk of missing relevant resource use, even if the respondent does not recognize the relevance, whereas capturing only disease-specific resource use would maximize causal precision and reduce the burden on respondents. We did not find any further recommendations.

3.4.2 Trial-Based Versus Model-Based Economic Evaluations

Economic evaluations can either be trial-based or model-based. Trial-based evaluations can be conducted alongside randomized controlled trials (RCTs). This phenomenon is called ‘piggybacking’. Studies agree that piggyback evaluations can be an appropriate means for conducting economic evaluations

(and employing RUM) [39, 57, 59, 60] provided methodological challenges are acknowledged. For example, when piggy-backing, often the main RCT outcomes determine the ideal sample size, while the ideal sample size for economic evaluation outcomes might be different [57]. No further information was found on RUM in model-based economic evaluations.

3.4.3 Psychometric Quality

All previously described aspects in the framework affect the overall quality of RUM in a study. However, it may be necessary to look at the measurement properties explicitly. Measurement properties refer to the validity and reliability of the resource use measurement instrument that is used for a study. For example, if RUM is conducted using a questionnaire, was this questionnaire validated beforehand, and how? Or was the questionnaire developed serving the aim of the study? One study reported that often only a subset of validated questions is used in a questionnaire; this would negatively influence the psychometric quality of the RUM instrument [4]. Nonetheless, if this is the case, it is recommended to have validated questions, at least for the main cost drivers. Other studies suggest using multiple data collection methods to increase the overall data quality [46, 61].

4 Discussion

Comprehensive guidance on valid RUM is lacking as previous studies have not been able to provide a general overview of methodological recommendations. There is a lack of transparency in the development process of existing RUM instruments, and current guidelines provide limited guidance about which aspects should be accounted for, and how, when conducting RUM. Therefore, the aims of this study were to (1) identify and cluster existing knowledge regarding RUM aspects, and (2) develop a framework that provides a comprehensive overview of methodological aspects of RUM in economic evaluations.

4.1 Methodological Reflection

To fulfil the aims of this study, an extensive literature search was conducted, synthesizing a sound rationale for appropriate RUM. As the clustering of available information generated new insights and showed the complexity of RUM, this study adds value to the existing literature, which has focused mostly only on addressing one RUM aspect at a time. Furthermore, the involvement of international health economic experts throughout the study phases has increased the validity of the framework. Nonetheless, the study was also prone to several limitations. As this was the first attempt to generate an overview of the existing evidence, no quality

assessment was performed for the included studies. It was decided that gathering the existing evidence outweighs the need for including studies based on quality assessment scores. In addition, to date no acknowledged quality criteria regarding methodological studies exist. Furthermore, the search strategy of the scoping review aimed for the retrieval of RUM aspects in general and did not focus on specific RUM aspects. Nevertheless, the probability of missing relevant studies is low as the approach towards identifying existing literature was thorough. Furthermore, even though experts were involved in all phases, the number of experts was limited and this might have affected the way that the RUM framework was constructed.

In addition, the results of this study revealed some challenges regarding the current status quo of RUM recommendations. First, in general, methodological evidence for all RUM aspects is scarce; only a few studies that intended to address RUM from a methodological perspective could be identified. Second, while four RUM domains were distinguished, the included articles focused mostly on the RUM domain ‘How to measure’ (29 of 40 included studies addressed this issue). Other RUM domains (e.g. ‘What to measure’ and ‘Additional considerations’) have received little to no attention. Third, published results can be contradictory; for example, studies disagreed on the level of reliability of self-reported resource use data in comparison with resource use data extracted from medical records [27, 28]. As most studies did not specify the generalizability of their conclusions to different settings, nor did they recommend a specific approach as a gold standard, it remains difficult to compare the results thoroughly. These findings both highlight the need for strong evidence-based recommendations in RUM and explain the lack thereof.

Although RUM methods and instruments need to be adaptive to contextual factors to capture context-specific resource use, insights from these studies can be used to increase the validity of RUM. The results of this study can help researchers to enhance methodological research on RUM development as the results address the importance and possible consequences of different RUM methods. While acknowledging the importance of external factors such as study design and national guidelines in selecting perspectives, this study adds value as the results facilitate more evidence-based choices that researchers can take into account before the start of the study. The results of this study can also be used to broaden the scope of existing national HTA guidelines to provide more structured guidance on the aspects that should be considered for adequate RUM in health economics research. In addition, existing databases such as DIRUM can use the framework to catalogue the existing research on each aspect of each RUM domain. Furthermore, the clustered information per RUM domain helps put existing RUM methodologies in perspective. In addition to applied health economic

studies, the results of this study also shed light on a possible focus for future methodological studies. To decrease the existing knowledge gap in RUM, future studies could focus on one single aspect of the framework and explore its validity, e.g. assessing the validity of proxy responses. In general, both applied and methodological research is needed to enhance the methodological body of RUM in economic evaluations, and inherently for the development of concrete RUM guidelines. Future studies could therefore focus on the development of a checklist for sound RUM decision making, for example one comparable with the checklist for judging preference-based measures by Brazier et al. [62].

5 Conclusion

The input of experts and existing literature regarding RUM recommendations were synthesized to develop a comprehensive framework to address the aims of this study. Its development was an iterative process and multiple health economics experts were involved in each phase. The final framework contains four RUM domains, each of which is further subdivided into more specific aspects that ought to be considered when deciding on, or interpreting the methodology of, the RUM approach. Existing methodological RUM findings extracted from the literature were clustered according to the framework. The results of the scoping review show the complexity of RUM; it encompasses a variety of aspects, some of which are interlinked, indicating that the choices for one RUM aspect may also affect other RUM aspects. For example, the ideal recall period (RUM Domain 3) depends on, among other factors, who is completing the RUM instrument (RUM Domain 1) and what resource use is asked. Consequently, others might argue for a different setup of the framework. While acknowledging that much has yet to be untangled, we believe that the current version is both useful and accurate in identifying the distinction between different domains and aspects.

The current study also provides valuable information for policy makers. The results of this study highlight the importance of appropriate RUM methodology, as it may directly or indirectly affect the outcome of a health economics study. Policy makers may use these findings to review the reliability of study results from a more evidence-based perspective.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s40273-021-01048-z>.

Declarations

Funding With the exception of Kirsty Garfield, all co-authors are partially funded by the PECUNIA project, which has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No. 779292.

Conflict of interest Luca Janssen, Ruben Drost, Aggie Paulus, Kirsty Garfield, William Hollingworth, Sian Noble, Joanna Thorn, Irina Pokhilenko, and Silvia Evers have no conflicts of interest to declare.

Availability of data and material All data generated or analyzed during this study are included in this published article and its supplementary information files.

Consent Not applicable.

Author contributions All authors contributed to the study conception and design. Material preparation and data collection and analysis were performed by LMMJ and the results were shared with all authors. The first draft of this manuscript was written by LMMJ, and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License, which permits any non-commercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc/4.0/>.

References

1. Buchanan J, Wordsworth S. Welfarism versus extra-welfarism: can the choice of economic evaluation approach impact on the adoption decisions recommended by economic evaluation studies? *Pharmacoeconomics*. 2015;33(6):571–9. <https://doi.org/10.1007/s40273-015-0261-3>.
2. Hakkaart-van Roijen L, Van der Linden N, Bouwmans C, Kanters T, Tan SS. Kostenhandleiding. Methodologie van kostenonderzoek en referentieprijzen voor economische evaluaties in de gezondheidszorg. In opdracht van Zorginstituut Nederland Geactualiseerde versie; 2015.
3. Noben CY, De Rijk A, Nijhuis F, Kottner J, Evers S. The exchangeability of self-reports and administrative health care resource use measurements: assessment of the methodological reporting quality. *J Clin Epidemiol*. 2016;74:93–106.
4. Leggett LE, Khadaroo RG, Holroyd-Leduc J, Lorenzetti DL, Hanson H, Wagg A, et al. Measuring resource utilization: a systematic review of validated self-reported questionnaires. *Medicine*. 2016;95(10): e2759. <https://doi.org/10.1097/md.0000000000002759>.
5. Drost RM, Paulus AT, Ruwaard D, Evers SM. Inter-sectoral costs and benefits of mental health prevention: towards a new classification scheme. *J Ment Health Policy Econ*. 2013;16(4):179–86.
6. PECUNIA. Vision and Mission. 2018. Available at: <https://pecunia-project.eu/project/vision-and-mission>. Accessed 30 Apr 2019.
7. Drummond M, Sculpher M, Torrance G, O'Brien B, Stoddart G. *Methods for the economic evaluation of health care programmes*. 2002.

8. ISPOR. Pharmacoeconomic Guidelines Comparative Table. 2019. Available at: <https://tools.ispor.org/peguidelines/COMP1.asp>. Accessed 18 Sep 2019.
9. Thorn JC. Resource-use measurement based on patient recall: issues and challenges for economic evaluation. *Appl Health Econ Health Policy*. 2013;11(3):155–61. <https://doi.org/10.1007/s40258-013-0022-4>.
10. Ridyard CH, Hughes DA, Team D. Development of a database of instruments for resource-use measurement: purpose, feasibility, and design. *Value Health*. 2012;15(5):650–5. <https://doi.org/10.1016/j.jval.2012.03.004>
11. Byford S, Leese M, Knapp M, Seivewright H, Cameron S, Jones V, et al. Comparison of alternative methods of collection of service use data for the economic evaluation of health care interventions. *Health Econ*. 2007;16(5):531–6. <https://doi.org/10.1002/hec.1175>.
12. EUnetHTA Joint Action 2, Work Package 7, Subgroup 3; Heintz E, Gerber-Grote A, Ghabri S, Hamers FF, Rupel VP, et al. Is there a European view on health economic evaluations? Results from a synopsis of methodological guidelines used in the EUnetHTA Partner Countries. *Pharmacoeconomics*. 2016;34(1):59–76.
13. The AMCP format for formulary submissions: welcome to Version 4.0. *J Manag Care Spec Pharm*. 2016;22(5):444–6. <https://doi.org/10.18553/jmcp.2016.16071>.
14. National Institute for Health and Care Excellence. NICE Process and Methods Guides. Guide to the Methods of Technology Appraisal 2013. London: National Institute for Health and Care Excellence; 2013. Copyright © 2013 National Institute for Health and Clinical Excellence, unless otherwise stated. All rights reserved; 2013.
15. Langley P. Dreamtime: Version 5.0 of the Australian Guidelines for Preparing Submissions to the Pharmaceutical Benefits Advisory Committee (PBAC). *Innov Pharm*. 2017. <https://doi.org/10.24926/21550417.1346>.
16. Clark RE, Teague GB, Ricketts SK, Bush PW, Keller AM, Zubkoff M, et al. Measuring resource use in economic evaluations: determining the social costs of mental illness. *J Ment Health Adm*. 1994;21(1):32–41. <https://doi.org/10.1007/bf02521343>.
17. Roberts RO, Bergstralh EJ, Schmidt L, Jacobsen SJ. Comparison of self-reported and medical record health care utilization measures. *J Clin Epidemiol*. 1996;49(9):989–95. [https://doi.org/10.1016/0895-4356\(96\)00143-6](https://doi.org/10.1016/0895-4356(96)00143-6).
18. Korthuis PT, Asch S, Mancewicz M, Shapiro MF, Mathews WC, Cunningham WE, et al. Measuring medication: do interviews agree with medical record and pharmacy data? *Med Care*. 2002;40(12):1270–82. <https://doi.org/10.1097/01.mlr.0000036410.86742.27>.
19. van den Brink M, van den Hout WB, Stiggelbout AM, van de Velde CJ, Kievit J. Cost measurement in economic evaluations of health care: whom to ask? *Med Care*. 2004;42(8):740–6.
20. Noben CY, Evers SM, Nijhuis FJ, de Rijk AE. Quality appraisal of generic self-reported instruments measuring health-related productivity changes: a systematic review. *BMC Public Health*. 2014;14:115. <https://doi.org/10.1186/1471-2458-14-115>.
21. Mayer S, Paulus ATG, Laszewska A, Simon J, Drost R, Ruwaard D, et al. Health-related resource-use measurement instruments for intersectoral costs and benefits in the education and criminal justice sectors. *Pharmacoeconomics*. 2017;35(9):895–908. <https://doi.org/10.1007/s40273-017-0522-4>.
22. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. *Int J Soc Res Methodol Theory Pract*. 2005;8(1):19–32. <https://doi.org/10.1080/1364557032000119616>.
23. Mays N, Roberts E, Popay J. Synthesising research evidence. London: Routledge; 2001. p. 194.
24. DIRUM. Database of Instruments for Resource Use Measurement. Available at: <http://www.dirum.org/>. Accessed 15 Jan 2020.
25. Moher D, Liberati A, Tetzlaff J, Altman DG, The PG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med*. 2009;6(7): e1000097. <https://doi.org/10.1371/journal.pmed.1000097>.
26. Green S, Kaufert J, Corkhill R, Creese A, Dunt D. The collection of service utilisation data: a research note on validity. *Soc Sci Med Part A Med Psychol Med Sociol*. 1979;13:231–4. [https://doi.org/10.1016/0271-7123\(79\)90032-4](https://doi.org/10.1016/0271-7123(79)90032-4).
27. Ritter PL, Stewart AL, Kaymaz H, Sobel DS, Block DA, Lorig KR. Self-reports of health care utilization compared to provider records. *J Clin Epidemiol*. 2001;54(2):136–41. [https://doi.org/10.1016/S0895-4356\(00\)00261-4](https://doi.org/10.1016/S0895-4356(00)00261-4).
28. Bellón JA, Lardelli P, Luna JD, Delgado A. Validity of self reported utilisation of primary health care services in an urban population in Spain. *J Epidemiol Community Health*. 2000;54(7):544–51. <https://doi.org/10.1136/jech.54.7.544>.
29. Cheung AH, Dewa CS, Wasylenki D. Economic grand rounds: impact on cost estimates of differences in reports of service use among clients, caseworkers, and hospital records. *Psychiatr Serv*. 2003;54(10):1328–30. <https://doi.org/10.1176/appi.ps.54.10.1328>.
30. Chung S, Domino ME, Jackson EW, Morrissey JP. Reliability of self-reported health service use: evidence from the women with co-occurring disorders, and violence study. *J Behav Health Serv Res*. 2008;35(3):265–78. <https://doi.org/10.1007/s11414-007-9105-z>.
31. Richards SH, Coast J, Peters TJ. Patient-reported use of health service resources compared with information from health providers. *Health Soc Care Community*. 2003;11(6):510–8.
32. Breda CS. Parent and institutional agreement on children's use of mental health services. *Eval Program Plann*. 1996;19(2):165–73. [https://doi.org/10.1016/0149-7189\(96\)00007-9](https://doi.org/10.1016/0149-7189(96)00007-9).
33. Hoogendoorn M, van Wetering CR, Schols AM, Rutten-van Mölken MPMH. Self-report versus care provider registration of healthcare utilization: Impact on cost and cost-utility. *Int J Technol Assess Health Care*. 2009;25(4):588–95. <https://doi.org/10.1017/S0266462309990432>.
34. Kashner TM, Suppes T, Rush AJ, Altshuler KZ. Measuring use of outpatient care among mentally ill individuals: a comparison of self reports and provider records. *Eval Program Plann*. 1999;22(1):31–40. [https://doi.org/10.1016/S0149-7189\(98\)00038-X](https://doi.org/10.1016/S0149-7189(98)00038-X).
35. Mistry H, Buxton M, Longworth L, Chatwin J, Peveler R. Comparison of general practitioner records and patient self-report questionnaires for estimation of costs. *Eur J Health Econ*. 2005;6(3):261–6. <https://doi.org/10.1007/s10198-005-0300-7>.
36. D'Souza-Vazirani D, Minkovitz CS, Strobino DM. Validity of maternal report of acute health care use for children younger than 3 years. *Arch Pediatr Adolesc Med*. 2005;159(2):167–72. <https://doi.org/10.1001/archpedi.159.2.167>.
37. Tisnado DM, Adams JL, Liu H, Damberg CL, Chen WP, Hu FA, et al. What is the concordance between the medical record and patient self-report as data sources for ambulatory care? *Med Care*. 2006;44(2):132–40.
38. Bresnahan BW, Rundell SD. Including patient-reported outcomes and patient-reported resource-use questionnaires in studies. *Acad Radiol*. 2014;21(9):1129–37.
39. Oostenbrink JB, Koopmanschap MA, Rutten FFH. Standardisation of costs—the Dutch manual for costing in economic evaluations. *Pharmacoeconomics*. 2002;20(7):443–54. <https://doi.org/10.2165/00019053-200220070-00002>.
40. Simoens S. Health economic assessment: a methodological primer. *Int J Environ Res Public Health*. 2009;6(12):2950–66. <https://doi.org/10.3390/ijerph6122950>.
41. Petrou S, Murray L, Cooper P, Davidson LL. The accuracy of self-reported healthcare resource utilization in health economic

- studies. *Int J Technol Assess Health Care*. 2002;18(3):705–10. <https://doi.org/10.1017/S026646230200051X>.
42. Short ME, Goetzel RZ, Pei X, Tabrizi MJ, Ozminkowski RJ, Gibson TB, et al. How accurate are self-reports? Analysis of self-reported health care utilization and absence when compared with administrative data. *J Occup Environ Med*. 2009;51(7):786–96. <https://doi.org/10.1097/JOM.0b013e3181a86671>.
43. Wolinsky FD, Miller TR, An H, Geweke JF, Wallace RB, Wright KB, et al. Hospital episodes and physician visits: the concordance between self-reports and medicare claims. *Med Care*. 2007;45(4):300–7. <https://doi.org/10.1097/01.mlr.0000254576.26353.09>.
44. Bowling A. Mode of questionnaire administration can have serious effects on data quality. *J Public Health*. 2005;27(3):281–91. <https://doi.org/10.1093/pubmed/fdi031>.
45. Robling MR, Ingledew DK, Greene G, Sayers A, Shaw C, Sander L, et al. Applying an extended theoretical framework for data collection mode to health services research. *BMC Health Serv Res*. 2010;10(1):180. <https://doi.org/10.1186/1472-6963-10-180>.
46. Schiotz M, Bogelund M, Willaing I. Challenges using online surveys in a Danish population of people with type 2 diabetes. *Chronic Illn*. 2012;8(1):56–63.
47. Ridyard CH, Hughes DA, Team D. Taxonomy for methods of resource use measurement. *Health Econ*. 2015;24(3):372–8. <https://doi.org/10.1002/hec.3029>.
48. Littvay L. Questionnaire design considerations with planned missing data. *Rev Psychol*. 2009;16(2):103–13.
49. Marques E, Johnson EC, Gooberman-Hill R, Blom AW, Noble S. Using resource use logs to reduce the amount of missing data in economic evaluations alongside trials. *Value Health*. 2013;16(1):195–201. <https://doi.org/10.1016/j.jval.2012.09.008>.
50. Noble SM, Hollingworth W, Tilling K. Missing data in trial-based cost-effectiveness analysis: the current state of play. *Health Econ*. 2012;21(2):187–200. <https://doi.org/10.1002/hec.1693>.
51. Clarke PM, Fiebig DG, Gerdtham U-G. Optimal recall length in survey design. *J Health Econ*. 2008;27(5):1275–84. <https://doi.org/10.1016/j.jhealeco.2008.05.012>.
52. Zhang W, Bansback N, Anis AH. Measuring and valuing productivity loss due to poor health: a critical review. *Soc Sci Med*. 2011;72(2):185–92.
53. Kjellsson G, Clarke P, Gerdtham UG. Forgetting to remember or remembering to forget: a study of the recall period length in health care survey questions. *J Health Econ*. 2014;35:34–46. <https://doi.org/10.1016/j.jhealeco.2014.01.007>.
54. Chou SL, Lamoureux E, Keeffe J. Methods for measuring personal costs associated with vision impairment. *Ophthalmic Epidemiol*. 2006;13(6):355–63.
55. Mirandola M, Bisoffi G, Bonizzato P, Amadeo F. Collecting psychiatric resources utilisation data to calculate costs of care: a comparison between a service receipt interview and a case register. *Soc Psychiatry Psychiatr Epidemiol*. 1999;34(10):541–7. <https://doi.org/10.1007/s001270050173>.
56. McCrone P. Mental health economics: current methodological issues. *Epidemiol Psychiatr Sci*. 2011;20(3):239–43.
57. Thorn JC, Noble SM, Hollingworth W. Methodological developments in randomized controlled trial-based economic evaluations. *Expert Rev Pharmacoecon Outcomes Res*. 2014;14(6):843–56. <https://doi.org/10.1586/14737167.2014.953934>.
58. Seidl H, Meisinger C, Wende R, Holle R. Empirical analysis shows reduced cost data collection may be an efficient method in economic clinical trials. *BMC Health Serv Res*. 2012;12:318. <https://doi.org/10.1186/1472-6963-12-318>.
59. O'Sullivan AK, Thompson D, Drummond MF. Collection of health-economic data alongside clinical trials: is there a future for piggyback evaluations? *Value Health*. 2005;8(1):67–79. <https://doi.org/10.1111/j.1524-4733.2005.03065.x>.
60. Coyle D, Lee K, Drummond M. Comparison of alternative sources of data on health service encounters. *J Health Serv Res Policy*. 1999;4(4):210–4. <https://doi.org/10.1177/135581969900400404>.
61. Evans CJ, Crawford B. Data collection methods in prospective economic evaluations: how accurate are the results? *Value Health*. 2000;3(4):277–86.
62. Brazier JE, Rowen D, Mavranzeouli I, Tsuchiya A, Young T, Yang Y, et al. Developing and testing methods for deriving preference based measures of health from condition-specific measures (and other patient-based measures of outcome). *Health Technol Assess*. 2012;16(32):1–109.